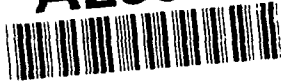


AD-A239 220



TATION PAGE

Form Approved
GAS No. 0704-0188

2

To average 1 hour per volume, including the time for reviewing instructions, searching existing data sources, gathering and reviewing the collection of information, sending comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

| | | | | | |
|---|--|---|--|---|--|
| 1. AGENCY USE ONLY (Leave blank) | | 2. REPORT DATE | | 3. REPORT TYPE AND DATES COVERED FINAL 01 Apr 89 to 30 Oct 90 | |
| 4. TITLE AND SUBTITLE PROBABILITY AND STATISTICS APPLIED TO THE THEORY OF ALGORITHMS | | | | 5. FUNDING NUMBERS AFOSR-89-0301 61102F/2304A5 | |
| 6. AUTHOR(S) J. Michael Steele | | | | 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) THE TRUSTEES OF PRINCETON UNIVERSITY OFFICE OF RESEARCH AND PROJECT ADMINISTRATION PRINCETON, NJ 08544-0636 | |
| 8. PERFORMING ORGANIZATION REPORT NUMBER AFOSR-89-0301 | | | | 9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/RI Bldg 410 Bolling AFB DC 20332-6448 | |
| 10. SPONSORING / MONITORING AGENCY REPORT NUMBER AFOSR-89-0301 | | | | 11. SUPPLEMENTARY NOTES | |
| 12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited. | | | | 12b. DISTRIBUTION CODE DTIC SELECTED S AUG 07 1991 D | |
| 13. ABSTRACT (Maximum 200 words) This report outlines in three sections the progress that has been made in the last two years. Work on the minimal spanning tree problem is first discussed, since this area has seen the most striking progress. The second section discuss work on the convex hulls of random walks. This work is the most recent, and it illustrates the broader applicability of ideas that were developed in the earliest stages of this grant. The third section discusses the theis work of doctoral candidates Maki Monna and Jun Gao who have been supported in part through this contract. | | | | | |
| 14. SUBJECT TERMS | | | | 15. NUMBER OF PAGES | |
| 17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED | | | | 16. PRICE CODE | |
| 18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED | | 19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED | | 20. LIMITATION OF ABSTRACT SAR | |

AFOSR-89-0301
Probability and Statistics Applied
to the Theory of Algorithms
Principal Investigator: J. Michael Steele

This report outlines in three sections the progress that has been made in the last two years. Work on the minimal spanning tree problem is first discussed, since this area has seen the most striking progress. The second section discuss work on the convex hulls of random walks. This work is the most recent, and it illustrates the broader applicability of ideas that were developed in the earliest stages of this grant. The third section discusses the thesis work of doctoral candidates Maki Momma and Jun Gao who have been supported in part through this contact.

I. *Solution of Bland's Conjecture on the Edges of the MST.*

Almost five years ago Professor Robert Bland (Cornell University, Operations Research) discovered through computer simulation that the minimum spanning tree (MST) of a random sample of n points from the unit square apparently has the property that the sum of the squares of the lengths of the edges must converge to a constant as n tends to infinity.

There are several factors that contribute to the importance of this observation. In the first place, MST's are among the most important objects of study in the field of computational geometry. They have basic relations to other key structures (like the Delaunay triangulation) and to algorithms (like Christophides' heuristic for the traveling salesman problem (TSP)). Second, MST's provide the most important example of a subadditive Euclidean functional that fails to be montone. This later feature is at the source of the difficulty of the well studied Steiner Problem (MST's where you can add additional points if they help). More important from the viewpoint of applied probability is that the non-monotonicity of the MST provides a basic challenge to the tools of subadditive Euclidean functionals. These methods are among the most systematic and general that have been developed to deal with asymptotics of problems in computational geometry and combinatorial optimization.

In Steele(1988, *Annals of Probability*) the first analysis is given for asymptotic behavior of the MST of random samples. In particular, it is proved there that we have

$$\sum_e |e|^\alpha \sim c_\alpha n^{(d-\alpha)/d} \int_{\mathbb{R}^d} f(x)^{(d-\alpha)/d} dx.$$

as $n \rightarrow \infty$ with probability one for all $1 \leq \alpha < d$ whenever the sample distribution determined by $f(x)dx + d\mu_s$ has compact support in \mathbb{R}^d . Here $f(x)$ is the density of the absolutely continuous part of the distribution, and μ_s is the measure corresponding to the singular part.

This result required a sustained analysis that usefully extended the theory of subadditive Euclidean functionals, and the techniques developed there are applicable to many other functions of interest in combinatorial optimization (e.g. the minimal matching problem and the semi-matching problem). Still, the approach of Steele(1988) falls short of resolving Bland's conjecture because it hinges critically the restriction that $\alpha < d$.

The next step toward Bland's conjecture was taken in Steele(1990), a paper that deals with many of themes that have developed out of the seminal work of Beardwood, Halton, and Hammersley on the TSP. In the course of other work, a proof is given there that the sum considered by Bland

91-06960


91 8 05 007

is bounded independently of n with probability one. This fact, proved via the spacefilling heuristic, added credibility and intensity to the conjecture.

The final proof of Bland's conjecture is obtained in Aldous and Steele(1990). The technique developed there differs hugely from those that have traditionally succeeded in combinatorial optimization. The key idea is to find an infinite object that is — in a rigorous sense — the limit of the MST of a random sample. Naive approaches to such an object fail miserably (e.g. the tree constructed for the Poisson process starting at a given point cannot be proved to span unless one assumes an unknown hypothesis from continuum percolation), but once a suitable infinite limit object is finally diagnosed, one has at hand a very powerful tool, the weak convergence theory of point processes. With Bland's conjecture cast in that framework, progress follows quickly. A side benefit of the proof of Bland's conjecture given in Aldous and Steele(1990) is that the weak convergence theory of point processes (which is sometimes held to be isolated and abstract) is demonstrated to be of real service on a concrete and resistant problem.

The work described above has been presented at several seminars, including Yale and the Institute for Defense Analysis(in Princeton). It was also presented as part of the recent SIAM Applied Probability Minisymposium in *Probability Applied to the Theory of Algorithms* organized by the PI.

II. Convex Hulls of Random Walks.

In collaboration with T.L. Snyder (Department of Computer Science, Georgetown University), recent progress has been made on the geometry of the convex hull of a random walk in \mathbb{R}^2 .

The problem of computing the convex hull is one of the central problems in computational geometry, and, although the issues seem well in hand in \mathbb{R}^2 for random samples, tough issues remain in \mathbb{R}^d for $d > 2$ and for *structured* subsets of \mathbb{R}^2 . Structured sets have computational interest because they sometimes allow for faster algorithms than general set, but such speed-up possibilities do not always exist. Thus, the algorithmic theory of structured sets contributes to our understanding of computational complexity. The best known algorithms for structured sets are perhaps those for sorting *sum sets*. Random walks have a structure that is both subtler and simpler than pure random samples, so they are an interesting target on which to test the complexity of the convex hull problem.

Earlier work in probability theory is also basic to the appeal of convex hulls of random walks. If L_n denotes the length of the convex hull of the set $\{S_0, S_1, \dots, S_n\}$ of points on a random walk in \mathbb{R}^2 , a remarkably beautiful formula for the expectation of L_n was discovered by Spitzer and Widom:

$$EL_n = 2 \sum_{k=1}^n E|S_k|/k.$$

The technical challenge that remains after the work of Spitzer and Widom is to get detailed information on more than just the expectation. In Snyder and Steele(1990) the techniques that were useful in understanding the TSP (e.g. the Jackknife inequality of Steele(1986, *Annals of Statistics*) and the theory of martingale inequalities) were put to work on L_n . The simplest non-trivial result given in Steele and Snyder(1990) is that $Var L_n = O(n)$. This result already suffices to supplement the work of Spitzer and Widom with a Strong Law, but it is further developed to give a theory of large deviations.

III. Other Work including Thesis Research.

Central to the intention of the Grant is the support of graduate education, especially in the research phase of the Ph.d. The work reported below is not yet complete, but substantial progress has been made that should lead to the submission of several articles during the present year.

A. Thesis Research of Maki Momma.

Mrs. Momma will complete her dissertation at the end of this summer. Her work has focused on the distribution and sampling theory of *intercepted samples*. The simplest example of such sampling consists of looking at the age and eventual lifetime of (say) an electronic component that is currently in use, or of a computer job that is currently running. The fact that such an "intercepted" item is still alive prevents it from being used naively to obtain an unbiased estimate of the lifetime distribution (of an item from the population of all items that have been placed in service). This observation leads to a theory of length biased sampling distributions that offers many applications and insights. This theory has been developed most forcefully by Y. Vardi (formerly of Bell Laboratories, currently at Rutgers).

The heart of our understanding the performance of many heuristic methods comes via the interpretation of simulations, and one of the important issues being addressed by Mrs. Momma is the application of biased sampling models to simulations. The results are not yet definitive, but biased sampling model seems to contribute a deeper understanding of real-time routing in communication networks and of other real-time heuristics like those employed in the self-organizing list.

B. Thesis Research of Jun Gao.

The work of Mr. Gao addresses the large deviation behavior of Euclidean functionals like the length of the shortest tour through a random sample (the TSP). He is in his third year of graduate school, and he has a solid mastery of the literature relevant to his research. In particular, he has read the papers of Rhee and Talagrand, E. Shamir, B. Bollobas, A. Frieze, and C. McDiarmid. These works all deal with the use of bounded difference martingales in order to obtain sharp tail bounds on variables of importance in combinatorial optimization.

Mr. Gao's most concrete accomplishment so far has been to systematize and simplify many of the techniques used by the previous authors. His immediate goal is to see how one can go beyond what is possible using bounded differences by using the martingale inequalities developed by Burkholder and others. The indications that this is a fruitful direction are outlined in Steele(1990, Cambridge University Press).

C. Further Research Activity.

During 1989 the PI published eight articles. Most of these articles deal with probability and the theory of algorithms. The main theme that is not mentioned above concerns asymptotic understanding of worst case problems, see items [3],[8], and [9] in the list of publications. While this activity seems to ignore probability, the details say otherwise, and one finds the interesting situation where just having a "large enough" problem guarantees a regularity analogous to that discovered in stochastic problems.

One paper has appeared so far in 1990, and five have been accepted and scheduled for publication. Two papers are under submission.

Grant Related Professional Activities

J. Michael Steele

The research activity supported by AFOSR-89-0301 is closely involved with three other professional activities of the PI. These activities and their relation to the supported research are outlined below.

I. *National Academy of Sciences: Board on Mathematical Sciences. Committee Report on Cross-Disciplinary Research.*

The PI currently serves as chair of a NAS-BMS Committee charged with developing a report "Probability and the Theory of Algorithms". This report will be written by a panel of eight experts in the field. So far the committee has written its research proposal and the proposal has passed through the review cycle of the BMS and NAS.

If the proposal is funded, the pannel will produce a report of approximately 100 pages that will review the "state of the art".

II. *Institute for Mathematics and Its Applications: Special Year in Applied Probability.*

The PI currently serves as chair of a committee constituted by the Board of the IMA to organize a Special Year in Applied Probability at the IMA in Minnesota for 1993-94. The year's activities will include about ten Workshops in Applied Probability, six to ten post-doctoral visitors to the IMA, about six long-term senior visitors, and several dozen short-term visitors. The year long activity may well provide the biggest single push that has been experience in the field of Applied Probability. It offers the opportunity to have an impact that will be felt for many years.

III. *Annals of Applied Probability.*

The PI currently serves as Editor of *Annals of Applied Probability*, the most recently created journal of the Institute of Mathematical Statistics. After years of review and discussion, the IMS acted to form this addition to the successful *Annals of Probability* and *Annals of Statistics*.

The new journal is off to a very promising start. A talented and energetic Board of Associate has been assembled. An editorial office with procedures following those of the earlier *Annals* has been equipped and organized. The news of the new journal has been published widely. Most important, the new *Annals* now receives steady flow of quality submissions. The first issue to appear in February 1991 is almost ready to be sent to the typesetters.

One of the charges of the new *Annals* was to facilitate IMS publication in the area of probability and computer science. From the looks of the first issue and the articles under review, the mission will be accomplished.